

Actual Causes as Changes of State in Continuous Time

Anonymous CogSci submission

Abstract

Keywords:

Introduction

Identifying the causes of an event involves not only logical rules but also different senses of “causation” that people might have. Let’s imagine that, to mitigate false alarms rate and improve the detection of fires, Lilly decides to install at home a dual alarm system that goes off if and only if two chains of ionization and photoelectric detectors are triggered (Fig.??a). Lilly wants to test her new system by exposing the detectors to some smoke. At some point a ionization detector starts beeping first while measuring a change in the electrical conductivity of air. It activates the next detector in the sequence which then activates the last one (Fig.??b). At this stage Lilly notices that the alarm remains silent. After a while a photoelectric detector starts beeping as well while measuring a change in the transparency of air. It activates the next detector in the sequence which then activates the last one and right after the alarm goes off (Fig.??c).

What would Lilly think caused the alarm to go off? Several answers can be considered. Based on logical dependencies, people’s causal judgements can state that both the furthest activated ionization detector (first root change “R1”) and the furthest activated photoelectric detector (second root change “R2”) are equally the causes of the alarm activation. However a more intuitive answer would be that the provided temporal information influences people’s choice toward one cause more than the other one. An important selection criterium could be how recent a candidate cause is compared to another one and would lead for instance to the following preferences: $R2 \succ R1$, and $I2 \succ I1$ (“I2” and “I1” being respectively the second and first immediate changes). Nonetheless we can imagine other cases where people potentially rely in part on another type of temporal criterium: let’s imagine that the activation of the ionization detectors starts before and finishes after the activation of the photoelectric detectors, such that R1 and I2 are linked by a *continuous sequence of changes*. In that case this is the activation of the oldest root change R1 and not the most recent root change R2 that initiate a continuous sequence of changes until the activation of the alarm. In such situations we want to know what people would consider as the main cause of the system.

The alarm case can be seen as an example of “singular causation”¹ where we want to understand, in a singular context, what caused a particular event to occur at the time it did. Here specifically it instantiates a class of causal systems where an event is a common effect of distinct sequences of necessary causes. Causal judgements have been widely studied in the literature, however the temporal order in which each variable changes its value is an important feature of causal systems that have been hardly studied nor tested experimentally as we intended to do. Indeed current research on actual causation mainly relies on two different interpretations of causality that don’t ground causal judgements on temporal information: the counterfactual (CF) and the physical process (PP) accounts. On the one hand, in the CF framework, events are represented as mere propositions and causation is thought as a relation between static states (A is a cause of B if and only if A and B are true, and A hadn’t occurred B wouldn’t have occurred). On the other hand the PP account is meant to give an epistemological definition of the concept of causation and not an explanation of how people actually rely on temporal informations to make causal judgements.

In this paper we present two experiments that show the importance of taking into account temporal information to explain the causal judgements people make. In particular we relied on the intuition behind the example above that the chronological order and the temporal relationship between distinct causes can bias people toward an event rather than another in considering which one is the main cause of the effect.

The continuity account of actual causation

This section uses the smoke detector example in Figure 1a to introduce our theoretical approach to actual causation. Like physical process theory, we focus on processes unfolding in continuous time, but it will be convenient to divide up the dimension of time into a sequence of intervals brief enough that there is at most one event per interval. The effect to be explained is an event that happens within one of these intervals: for example, a smoke alarm’s change of state from inactive to active. Given this setup, our theory can be formulated using a procedure that starts with the effect and steps

¹Also called “token” or “actual causation” by opposition with “general causation” – following a distinction that has been frequently made in the literature.

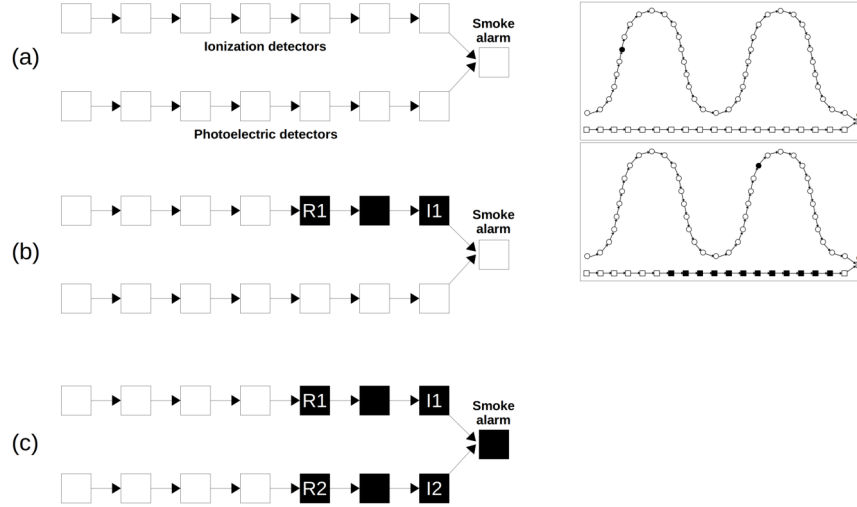


Figure 1: (a) An example of an AND-Gate causal system. In an AND-Gate for an effect to occur it needs the occurrence of all the causes it is linked to. Here for the smoke alarm to go off, both the ionization and photoelectric detectors it is connected to have to be activated. In this scenario the detectors activation spreads over the top chain first (ii) and then over the bottom chain leading to the alarm activation (iii).

backwards through time until the actual cause of the effect is identified.

The immediate cause of the smoke alarm’s activation must be an event that took place in the interval preceding the activation. In Figure 1a, the event immediately preceding the alarm’s activation was the activation of I_C . Having identified this immediate cause, we then step backwards and identify the immediate cause of this event, and so on. The procedure terminates once we arrive at an event that has no immediate cause within the system of interest, and this cause is declared to be the main cause of the effect. In Figure 1a the main cause is the activation of R_C .

In most cases, the backward-tracing procedure just described will identify a single main cause of the effect. If the effect has no immediate cause within the system, then no main cause will be identified. Because there is at most one event per interval, the procedure can never identify more than one main cause.

The assumption that at most one event occurs per interval follows from the idea that there are no coincidences. In continuous time it is exceedingly improbable that two events would occur at exactly the same time — in technical terms this kind of coincidence can be described as a “measure zero” possibility. If there are no coincidences, then slicing the time dimension sufficiently finely will ensure that there is at most one event per interval.

The boundaries of the causal system under consideration will depend on context. For simplicity, our discussion of Figure 1a focused only on events internal to the alarm system, and addressed the question of which detector is best viewed as the main cause of the alarm’s activation. Questions about smoke alarms, however, will often be framed with respect to

a causal system that includes both the alarm system and its surroundings. Expanding the causal system in this way can allow the alarm’s activation to be traced back to an event such as an electrical fault that started a fire which in turn triggered the alarm.

Experiments

We developed two experiments to test the continuity account of actual causation just presented. The cover story for both experiments asked participants to imagine that they worked in a control room and that their job was to monitor networks of particle detectors. Detectors were shown as white squares (as in Figure 1 that become active and turn black when they absorb a radioactive particle. Activation is transmitted across links in the network, and a detector activates if all of its input detectors are active. Each network contains a special detector called the Gauge of Critical Moment (GCM, analogous to the smoke alarm in Figure 1a) and participants were told that “for each activation sequence that you see, your job is to decide what caused the activation of the GCM.”

Both experiments included chains in which the GCM received input from one detector, and dual branch networks such as Figure 1a in which the GCM received input from two detectors. For chains, the continuity account predicts that participants will tend to choose the root cause of the effect (ie the detector that initiates the activation sequence) rather than the immediate cause (the detector that immediately precedes the GCM). For dual branch networks, the continuity account predicts that participants will tend to choose the root cause on the branch whose activation is temporally continuous with the activation of the GCM. We refer to this branch as the continuous branch, and refer to the root cause and the immediate cause

on this branch as R_C and I_C respectively, where the subscript denotes “continuous.” We refer to the other branch as the “delayed branch,” and use R_D and I_D for the root and immediate causes on this branch. The “delay” in this naming scheme refers to the delay that occurs between I_D and the activation of the GCM. It is convenient to use the same labels for both detectors and events: for example, R_C will be used to denote both a detector on the continuous branch and the activation of this detector. This flexible use of notation, however, should not be taken to suggest that actual causation involves a relationship between objects instead of a relationship between events.

Our presentation of the experiments focuses on two theoretical accounts: the continuity account and the probability raising account. For completeness, though, we first discuss the counterfactual and physical process accounts. Both accounts can be formulated in different ways — here we present what we take to be the most natural predictions of the two accounts, and the General Discussion considers ways in which the accounts can be adjusted to make different predictions.

The counterfactual account makes no clear prediction about whether root causes or immediate causes should take priority. For a chain network, if the root cause had not occurred, the effect would not have occurred, but it is equally true that if the immediate cause had not occurred, the effect would not have occurred. If we consider dual branch networks and restrict attention to root causes only, then the account makes no clear prediction about whether R_C should be preferred to R_D . If R_C had not occurred, then the effect would not have occurred, and likewise for R_D .

The physical process account is similarly inconclusive. Both root and immediate causes are connected by a physical process to the effect, and there seem to be no clear grounds for preferring one to the another. If we focus only on root causes, again there is no clear preference between R_C and R_D . After the activation of I_D , one can think of this detector as continuously sending activation towards the GCM, but this unbroken stream of activation only “unlocks” the detector once I_C is also active. It follows that both R_C and R_D are connected to the effect by physical processes.

Because both of these accounts make no clear predictions, it seemed possible that people’s inferences about the networks in our experiments would be highly variable and would reveal no clear trends. Our first experiment therefore explored simple cases analogous to the smoke detector example in Figure 1 with the goal of establishing whether the basic experimental procedure was viable. The second experiment focused on more elaborate dual-branch cases that aimed to distinguish between the continuity account and Spellman’s probability raising account (Spellman, 1997).

Experiment 1

Experiment 1 included both causal chains and dual branch networks. For all of the dual branch stimuli, I_D occurred before R_C (Figure ??a) which means that the delayed branch completed its activation before the continuous branch began

to activate. The probability raising and continuity accounts both predict that root causes should be preferred to immediate causes, and that for dual branch networks R_C should be preferred to R_D . The primary purpose of the experiment was to test both predictions.

Participants. 30 participants were recruited via Amazon Mechanical Turk and paid XXX for a YYY minute experiment.

Materials. The experiment used a customized interface built using the jsPsych library (?, ?). For all networks presented, participants clicked a “Run” button to observe an activation sequence. The first event in the sequence (i.e. the first change of state) always took place 5 seconds after the Run button was clicked, and the delay between successive activations along a chain of detectors was set to 100 ms. After the final event in a sequence (i.e. the activation of the GCM), all detectors became clickable after a delay of 1 second. Clicking on a detector turned its border red, and at most one detector could be selected at any time. After a sequence completed, participants could view it again if they wished by clicking a “Run again” button.

Design. The experiment included activation sequences over 15 networks (3 chains and 12 dual branch networks). Excluding the GCM, each chain and each branch of each dual network had 7 detectors. We refer to the sequences as *short*, *medium* or *long* based on the distance between the root causes and the GCM. Excluding the GCM, the short, medium and long chain sequences showed 1, 3 and 6 active detectors respectively at the end of the sequence. Within each dual branch sequence, the root causes on the two branches (R_D and R_C) were equidistant from the GCM, but these distances varied across sequences. The sequence in Figure 1a is a medium dual branch sequence (3 active detectors in each branch excluding the GCM), and short and long versions had 1 and 6 active detectors respectively per branch.

Each dual branch sequence (short, medium and long) came in four versions. The *state* version showed the delayed branch (including detectors R_D I_D) as active from the very beginning of the sequence. The activation of this branch was therefore presented as having occurred at some indefinite time in the past, resulting in a steady state of activation. The three *event* versions all showed R_D activating 5 seconds into the sequence, and had delays of 2, 4 and 6 seconds between the activation of I_D and R_C .

Procedure.

Participants first read a set of instructions which introduced the task and included examples of a chain and a dual branch network. They then answered three questions about the task and the detectors, and were sent back to read the instructions again if they answered incorrectly. They continued cycling through the instructions and the test questions until they answered all questions correctly.

The 15 activation sequences in the experiment proper were presented in random order. For dual branch sequences, the position of the delayed branch was also randomized (Figure 1

shows the delayed branch on the top rather than the bottom). The orientation of the network (GCM on the left or the right) was randomized within participants. The prompt after each sequence was “In this sequence what caused the activation of the GCM? Respond by clicking on a detector,” and participants were required to choose a single detector in response.

Results.

Network orientation (left or right) and position of the delay branch (top or bottom) had no apparent effect, and we therefore collapse across these variables. Figure ?? summarizes the results for short, medium and long dual-branch sequences. The delay between I_D and R_C had no significant effect, and Figure ?? combines results for all three delays.

Although R_D , I_D , R_C and I_C all qualify as causes of the effect, Figure ?? suggests that R_C tends to be singled out as the main cause. This result can be separated into two general conclusions. First, participants were more likely to choose root causes than immediate causes. Root and immediate causes were identical for the short sequences and therefore cannot be distinguished, but the results for medium and long sequences reveal a preference for root causes. STATISTICAL ANALYSIS. The data for chains (not shown in Figure ??) provide further evidence of a preference for root causes. NUMBERS HERE

The second conclusion is that for dual branch sequences, the root cause on the continuous branch (R_C) is preferred to the root cause on the delayed branch (R_D). This result is consistent with the continuity account but also consistent with the probability raising account, because R_C guarantees the occurrence of the effect but R_D does not.

Figure 1 suggests that event and state sequences lead to similar patterns of responses. The continuity account makes the same prediction for both kinds of sequences, but we thought that the event sequences might make participants especially likely to choose R_C over R_D . If this effect exists, our data suggest that it is relatively small.

It is also notable that people’s judgments are largely unaffected by the delay between I_D and R_C and the activation length of each sequence (short, medium or long). Both of these variables affect the time that elapses between the root causes and the effect, and our results suggest that people’s judgments are not exquisitely sensitive to this kind of variation.

Experiment 2

In the previous experiment we have seen that people generally chose R2 (identical to I2 in the short sequences) as the main the cause of the activation of the GCM. These results make sense since R2 and I2 are linked by a continuous sequence of changes. But this set up, where the first element of a chain “waits” for the occurrence of the last event of another chain before it can initiate a second sequence, is probably not the most common scenario in real life. For instance it might take more time for one causal process to end so that the causal process that has occurred first is not necessarily the one that ends first. In Experiment 2 we therefore wanted to

a) Sequence: R_D , I_D , R_C , I_C

	R_C	R_D
Counterfactual	✓	✓
Physical Process	✓	✓
Probability Raising	✓	
Continuity	✓	

b)

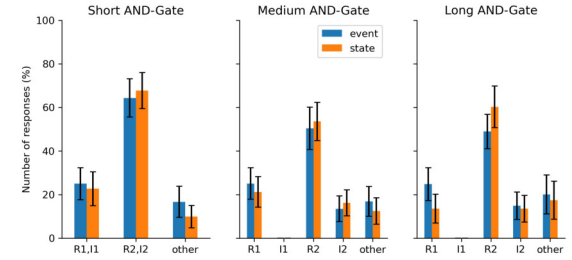


Figure 2: Experiment 1. Comparison of the results for different length of activation in AND-Gates. Short: 1 activated detector in each branch. Medium: 3 activated detectors in each branch. Long: 6 activated detectors in each branch. “State” responses correspond to the networks where one branch was already activated before participants run the animation. “Event” responses correspond to the networks where no detectors were activated before and include results for 2s, 4s and 6s delays between I1 and R2. BUT SEE WARNING ABOVE.

break the temporal continuity between the root changes and the immediate changes of Experiment 1 such that the continuous sequence of changes be sometimes between R1 and I2 or R2 and I2. We predicted that when there is a continuous sequence of changes between R1 and I2, people would preferably choose R1 as the main cause.

Participants. The experiment was hosted by Google App Engine and 98 participants were recruited via Amazon Mechanical Turk. There were 40 females (average age: 40.4), 57 males (average age: 37.6) and 1 declared “other”. All the participants were english speakers except for 2 participants.

Method. In Experiment 2 each participant was presented 7 different networks. One stimulus was a chain and all the other six stimuli were AND-Gates. Like in Experiment 1 we tested for the dimension “Event” vs “State”. Same as before in condition “Event” none of the detectors was initially active and in condition “State” one of the two branches was containing detectors that were already active before the animation started. However this time, within the condition “Event”, we also tested for the extra dimension “R1 cont” vs “R2 cont”. In condition “R1 cont” (represented in Fig.??) the activation is such that there is a continuous sequence of changes going from R1 to I2 leading to the activation of G. In condition “R2 cont” (not represented here) this is R2 that was continuously connected to I2 leading to the activation of G. For the AND-Gates the delay between I1 and I2 in condition “Event” was

maintained at 11 nodes (i.e. 1100 ms) for almost all the networks. There was only one network for which the delay was set at 8 nodes (800ms).

As in Experiment 1 participants were first presented the network in its initial state and were asked to click the “Run” button to observe an activation sequence over the network. Participants waited 5000ms to see a first change of state occurring in a detector. Activation delay between any two successive square activations was set on 100ms as well. All the stimuli were presented in a random order and participants were assigned to either group “Left” or “Right” as in Experiment 1, and each AND-Gate was randomly presented with the curved chain (see Fig.??) at the top or at the bottom.

Results. First we observed no significant effect (statistics?) of neither the orientation of the networks (left vs right) nor the location of the curved chain (top vs bottom). On average participants took 8 min to complete the experiment and overall 91% of the time participants run an animation only once. Second the reversed response pattern for R1 and R2 between “R1 cont”, “R2 cont” and “State” stimuli type shown in (Fig.3) confirms our hypothesis: when this is the first root change R1 that initiates a continuous sequence of changes until the activation of G, R1 is considered the cause of the activation of G about two times more than R2 (31% against 15%) while this is the opposite when this is the second root change that initiates a continuous sequence of changes (18% against 29%). Coherently, for “State” condition (when this is R2 that initiates the continuous sequence of changes) participants prefer R2 over R1 in a very large extent (47% against 5%). As expected the type of network had however no effect on the responses for I1 nor I2 and the percentages of responses for I2 and the root change that initiates the continuous sequence for the Event network (31% and 34% for respectively the “R1 cont” and “R2 cont” networks) were approximately the same.

a) R_C first

Sequence: R_C, R_D, I_D, I_C

	R_C	R_D
Probability Raising		✓
Continuity	✓	

b) R_C second

Sequence: R_D, I_D, R_C, I_C

Sequence: R_D, R_C, I_D, I_C

	R_C	R_D
Probability Raising	✓	
Continuity		✓

c)

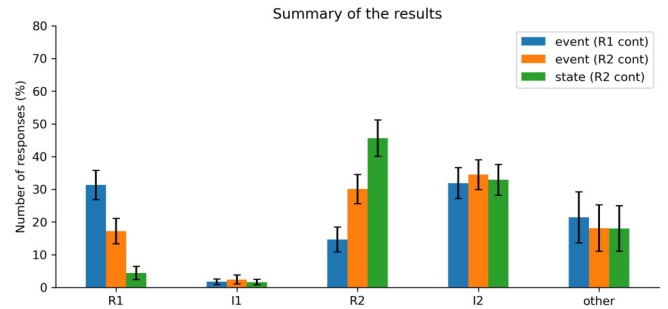


Figure 3: Results for AND-Gates where the continuous sequence of changes is either between R1 and I2 or between R2 and I2. “Event (R1 cont)”: responses for the networks where no detectors were already active before the participant runs the animation and where the continuous sequence of changes is between R1 and I2. “Event (R2 cont)”: responses for the networks where no detectors were already active and where the continuous sequence is between R2 and I2. “State (R2cont)”: responses for networks where one branch was already activated before the participant runs the animation and where the continuous sequence of activation is necessarily initiated by R2.

General discussion

Outline:

I. Summary and implications of the results

II. To say why these results cannot really be explained by the current theories of causation and why our theory can better account for them. I don't know if I should talk more about CF and PP accounts than I do in introduction, like saying something like that is probably not useful:

[According to the CF account, the general idea is that an event A is said to be a cause of a distinct event B if the occurrence of A makes a difference in the occurrence of B. More precisely A is a cause of B if and only if A and B are true, and A hadn't occurred B wouldn't have occurred. This interpretation calls upon counterfactual scenario or possible worlds where the presumed cause of the effect is removed from the system while all other relevant factors are kept as unchanged as possible compared to the actual world. According to us the main problem of this interpretation is that its best models explicitly rejects the idea of grounding causation on temporal information. According to the PP account, an event A causes B when there is a physical connection between them. In one of its latest and probably most convincing formulation, this theory distinguishes two things : causal process and causal interaction (i.e. causation). A causal process is a physical process involving an object which conserves a certain quantity, like mass-energy, accross space and time. A causal interaction or causation is an exchange of that conserved quantity. However the theory goes sometimes against some of our best causal intuitions and to makes implicitly use of the CF analysis to explain our judgements: if I hold the head of my ennemy under water and make him die, I'm not genuinely the cause of his death; rather I'm actually preventing the possibility of a genuine causation which is breathing oxygen in order to live.]

Or if I do have to explicit both theories I should insist on what the predictions of both theories are for our experiments and why they cannot explain people's intuition (the role of time).

III. To insist, maybe, on the work of Glymour where our project in part originated: explaining his caveat about treating events as pure propositions. To explain how our experiments extend this idea and insist more on the role of changes in continuous time. Maybe we should present some insights about how we could formalize further the model? How we could use formal tools like Graphical models?

IV. To present further experiments that would be interesting to make (with loops, cases of prevention, etc.)

References

Halpern, J. Y., & Pearl, J. (2001). Causes and explanations: A structural-model approach – Part I: Causes. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence* (p. 194-202). San Francisco, CA: Morgan Kaufmann.

Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General*, 126(4), 323–348.