[Preprint] What Does It Take To Make A Difference? A Reply To Andreas And Günther*

Sander Beckers†

Abstract

Andreas & Günther have recently proposed a difference-making definition of actual causation. In this paper I show that there exist conclusive counterexamples to their definition, by which I mean examples that are unacceptable to everyone, including AG. Concretely, I show that their definition allows c to cause e even when c is not a causal ancestor of e. I then proceed to identify their non-standard definition of causal models as the source of the problem, and argue that there is no viable strategy open to AG to fixing it. I conclude that their definition of causation is damaged beyond repair.

1 Introduction

Andreas & Günther (AG from now on) have recently proposed a difference-making definition of *actual causation*, which is the relation that takes place when one token event causes another event, such as when someone throws a rock and breaks a window (Andreas and Günther, 2021a). Similar to many

*This is a preprint version of a paper that is forthcoming in the *Journal of Philosophy*, under the "Comments and Criticism" section.

†University of Amsterdam, srekcebrednas@gmail.com, sanderbeckers.com

recent proposals, their definition is formulated using causal models and is built on a counterfactual condition. Contrary to existing proposals, AG offer their own non-standard definition of causal models. They do so because they want to formulate their counterfactual condition in terms of a model being *uninformative* on some variable e, which is not something that standard causal models allow.

The primary aim of this paper is to show that there exist conclusive counterexamples to their definition, meaning examples that are unacceptable to everyone, including AG. Concretely, I show that according to their definition, c can cause e *even when c is not a causal ancestor of e*. The secondary aim is to identify the source of the problem, and to argue that there is no viable strategy open to AG to fixing it. If they modify their notion of uninformativeness to avoid said counterexamples, novel conclusive counterexamples ensue. If they modify their definition of causal models to bring it closer to the standard definition, they end up with verdicts of causation that contradict what AG take to be the main advantage of their definition. I conclude that their definition is damaged beyond repair.

Elsewhere AG have proposed two other definitions of causation using causal models, but neither of them forms an alternative to the present definition, because neither belongs to Lewis (1973)'s counterfactual tradition that AG here explicitly espouse (Andreas and Günther, 2020, 2021b). (From now on Andreas and Günther (2020) is AG1, Andreas and Günther (2021a) is AG2, and Andreas and Günther (2021b) is AG3.) In addition, AG3 disagrees with AG2 on those verdicts that AG use precisely to motivate the superiority of AG2.

The paper proceeds as follows. The next section presents AG's general motivation for defining causation as AG2. Section 3 analyzes AG2 and its relation to other counterfactual definitions of causation. With this understanding in hand, Section 4 dives into the problems that plague AG2 and Section 5 explains why

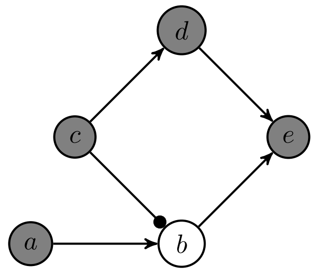

Figure 1: **Early Preemption**

they cannot be resolved by modifying AG's definitions, be it their definition of causal models or AG2 itself.

2 Motivating AG2

AG motivate AG2 by the claim that it is the only counterfactual-based definition on offer which arrives at the correct verdicts for two important examples, labelled **Early Preemption** and **Switch**, namely that $c = 1$ is a cause of $e = 1$ in **Early Preemption** but not in **Switch**. Here **Early Preemption** and **Switch** are nothing but the labels that AG – following Hall (2007) – give to the causal structures depicted in Figures 1 and 2, respectively. Strictly speaking AG's claim is false, for Hall's definition also arrives at these verdicts. However, as AG point out, Hitchcock (2009) has identified several fundamental problems with Hall's definition and hence they do not consider it.

The graphs in Figures 1 and 2 are examples of so-called "neuron diagrams", which are often used in the literature on causation to visually represent causal models. The nodes represent binary variables, and a node is gray iff the variable takes on the value 1. The edges represent causal influence in the following manner: the receiving node takes on the value 1 whenever any of the edges with an arrowhead comes from a node that has value 1 *and* there is no edge with a circlehead coming from a node that has value 1. For **Early Preemption** this means that the values of the variables are determined by the following equations: $c = 1$, $a = 1$, $d = c$, $b = \neg c \wedge a$, and lastly, $e = d \vee b$. For **Switch** the equations are: $c = 1$, $d = c$, $b = \neg c$, and $e = d \vee b$. We call the leftmost variables of a diagram *root variables*, as their values are not determined by other variables.

Note that in **Switch** the value of e always comes out as 1, regardless of the value of c. Furthermore, the upper structure is entirely symmetric to the lower one: either we have c leading to e via d, or we have $\neg c$ leading to e via b, all c does it to act as a "switch" that determines whether we follow the upper path or the lower path. It is because of this symmetry that AG claim c does not make enough of a difference to be considered a cause of e.

At first sight AG's claim is in agreement with the literature, which (for the most part) likewise believes that switches are not causes. However, there is substantial disagreement on whether or not **Switch** is an appropriate representation for the kind of informal story that characterizes switching behavior (Beckers and Vennekens, 2017; Halpern and Pearl, 2005). If it is not, then it does not speak against a definition that it treats **Switch** similarly to **Early Preemption** and calls c a cause of e in both.

In fact, treating **Switch** differently to **Early Preemption** is the source of one of the fundamental problems that Hitchcock (2009) identifies to criticize Hall (2007)'s definition. Concretely, he shows that by making what appears to be an innocuous modification to **Switch**, Hall's definition changes its verdict from c not being a cause to c being a cause.[1] Importantly, AG2 delivers the same verdicts for these examples. Given that, as mentioned earlier, AG explicitly invoke Hitchcock's criticism of Hall's definition, it is surprising that AG fail to notice this. Moreover, AG2 agrees with Hall's definition on *all* of the examples that AG discuss. Hence it is not at all clear why AG consider their definition an improvement over that of Hall's.

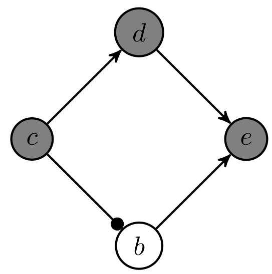

Figure 2: Switch

For the present purposes I set this discussion aside entirely, and simply accept AG's claim that c is not a cause of e in **Switch** whereas it is in **Early Preemption**.

3 Defining Causation using Causal Models

3.1 The Counterfactual Tradition

AG subscribe to the counterfactual tradition of defining actual causation, and they follow the recent trend within this tradition that embraces *causal models*, a rich formalism that enables a compact representation of the causal relations between variables (Halpern, 2016; Pearl, 2009). Causal models are widely used to define actual causation by filling in the following schema in various manners.

Definition 3.1: [**Actual Causation Schema**] c is a cause of e relative to causal model M iff

- (C1) M satisfies c and e, and
- (C2) there is a causal model M' that meets conditions Con such that $M'_{\{\neg c\}}$ satisfies $\neg e$.

Here $M'_{\{\neg c\}}$ is the causal model that results from M' after applying the intervention $\neg c$.

Filling in this schema requires two elements: a definition of causal models (including a definition of interventions), and a specification of the conditions Con. Existing definitions in the literature all use (essentially) the same standard definition of causal models and differ only with regards to the conditions Con that they impose.[2] Roughly speaking, the standard definition of a causal model M is simply the set of equations that describes the behavior of structures such as the neuron diagrams in **Switch** and **Early Preemption**, and the corresponding standard definition of an intervention that sets some variable x to a value i is the operation on M that replaces the existing equation for x with $x = i$.

The well-known recent definition of causation by Halpern (2016) is helpful to illustrate the second element. It specifies Con as the requirement that M' is the result of an intervention on M that holds fixed some variables at their actual values (meaning the values obtaining in M). Applied to **Early Preemption**, we see that c causes e by observing that if we hold fixed $\neg b$ and then intervene to get $\neg c$, the resulting equations tell us that $\neg e$ holds. The same reasoning applies to **Switch**, and hence Halpern's definition likewise considers c a cause of e there.

AG introduce a non-standard definition of causal models that allows the formulation of a condition Con in terms of a model being *uninformative* on c and e. This notion of uninformativeness is unique to their framework, and it is the key ingredient that allows them to distinguish **Switch** from **Early Preemption**.

3.2 The AG2 Definition

On AG's view a *causal model* $< M, V >$ consists of a set M of structural equations of the form $x = \varphi$ for some propositional formula φ and a consistent set of literals V. They also implicitly assume there to be a fixed set of propositional variables

P out of which the literals in V and the formulas φ can be constructed. If there is a single literal l for each variable p, then the set V describes a unique state of the world. Such a set is called *complete*. Although AG allow for a causal model to consist of a set V that is not complete, unless otherwise noted we assume that V is complete. The structural equations describe how the values of some variables in P are determined by the values of other variables. AG implicitly assume that for each variable in P there is at most one equation.

The semantics of a causal model are given by combining M with V in the following manner. If the truth values of V are consistent with all elements of M, then V is said to *satisfy* M. $< M, V >$ *satisfies a formula* φ iff φ is true in V or V does not satisfy M. In case V is not complete, satisfaction is defined by requiring satisfaction in all consistent complete extensions V^c of V that satisfy M. By allowing for V not to be complete, AG can introduce the idea that a causal model is *uninformative* regarding some formula φ, namely as the condition that the model satisfies none of φ and $\neg\varphi$.

An intervention I is a consistent set of literals that functions as an operator on causal models. However, rather confusingly, AG define interventions as an operator solely on the equations M of a causal model and in this context refer to M itself as the causal model, thus ignoring V altogether. Concretely, they define the result of an intervention I on M as $M_I = \{(p = \varphi) \in M | p \notin I \text{ and } \neg p \notin I\} \cup I$. When defining satisfaction of a formula in an intervened model, AG revert back to interpreting a model as a pair $< M_I, V >$ and state that satisfaction is simply defined identically as before.

This definition of an intervention needs to be slightly modified to make any sense. Say we have a trivial model such as $< \{c = 1\}, \{c\} >$. On the current definition, intervening to set c to 0 results in $< \{c = 0\}, \{c\} >$. Here V does not satisfy M, and thus we get logical explosion: $< \{c = 0\}, \{c\} >$ satisfies all formulas. Clearly that is an undesirable way for interventions to behave. Instead, we would expect that by intervening on c we also remove any literal containing c from V. Thus, I propose to define an intervention I as an operator on the *entire* model $< M, V >$, giving $< M_I, V_I >$, where M_I is as before and $V_I = \{l \in V | l \notin I\}$. This is a simple and unproblematic modification of AG's definition of an intervention. (Concretely, as will become clear, it does not affect their definition of causation.)

The standard definition of a causal model can be retrieved as the special case in which $V = \varnothing$ and M contains an equation for all variables. A standard causal model induces a unique complete assignment V^c: the values of the root variables are given by their constant equations, and these values can be filled in recursively into the equations of the other variables to determine their values.[3] (For example, in **Switch** all values are determined by the value of c, whereas in **Early Preemption** everything is determined by the values of both c and a.) This implies that a standard causal model is informative regarding all formulas.

The fact that AG's definition of a causal model *does not* require M to contain an equation for each variable explains why on their account a model can be uninformative regarding some formulas. AG2 builds on this property by spelling out the condition *Con* as being uninformative regarding the putative cause and effect, as follows. (AG also present a more general definition that allows for complex causes and effects, but for our discussion this definition suffices.)

Definition 3.2: [**Actual Causation**] Let $< M, V >$ be a causal model such that V satisfies M. c is a cause of e relative to $< M, V >$ iff

- (C1) $< M, V >$ satisfies c and e, and
- (C2) there is a $V' \subset V$ such that $< M, V' >$ is uninformative on c and e, while $< M_{\{\neg c\}}, V'_{\neg c} >$ satisfies $\neg e$.

Recall that $< M, V' >$ being uninformative on c and e means that $< M, V' >$ satisfies none of c, $\neg c$, e, and $\neg e$.

(Note that my modification to the definition of an intervention is irrelevant to Definition 3.2, since satisfying (C2) requires that $V'_{\neg c} = V'$ anyway.)

Before applying AG2 to the two examples, I should point out that there are several ways of representing neuron diagrams as a causal model under AG's definition. Firstly, for each root variable x we can choose to store its value in M, namely as an equation of the form $x = i$ (where i is 1 or 0), or we can choose to store its value in V, namely as a literal (x or $\neg x$, respectively). Secondly, for each non-root variable we can choose to store its inferred value in V or not. The two simplest choices are to store all the root variables in M and let V be empty – call this the *M-representation* – or to not store the root variables in M and let V be complete – call this the *V-representation*. For example, the M-representation of **Early Preemption** is $< \{c = 1, a = 1, d = c, b = \neg c \wedge a, e = d \vee b\}, \varnothing >$, whereas its V-representation is $< \{d = c, b = \neg c \wedge a, e = d \vee b\}, \{c, a, \neg b, d, e\} >$. Note that M-representations are standard causal models, and therefore they are informative on all formulas. Combine this with the fact that $V = \varnothing$ and we see that it makes little sense to apply AG2 to M-representations. For this reason AG implicitly follow the convention of always choosing the V-representation, i.e., they assume that M never contains equations of the form $x = i$ and V is complete.

Applying AG2 to **Switch**, it is easy to see why it fails to call c a cause of e: no matter whether we assume c or $\neg c$, the equations determine that e holds, and thus there does not exist any set V' so that $< M, V' >$ is uninformative regarding e.

Applying AG2 to **Early Preemption**, we can choose $V' = \{\neg b\}$, thereby "forgetting" that we started out with c and a. Given the equations, there exist three complete extensions of V': $V_1 = \{\neg b, c, d, e, a\}$, $V_2 = \{\neg b, c, d, e, \neg a\}$, and

$V_3 = \{\neg b, \neg c, \neg d, \neg e, \neg a\}$. Since all of c, $\neg c$, e, and $\neg e$ appear in at least one of these extensions, $< M, V' >$ meets the requirement that it is uninformative on c and e. Intervening with $\neg c$ means to add $c = 0$ to the equations, from which it follows that the only complete extension is V_3. As $\neg e \in V_3$, we get that c causes e, as desired.

4 Counterexamples to AG2

4.1 Counterfactual Dependence

Before presenting the more fundamental problems with AG2, we warm up with a problem that *can* be fixed. It can best be highlighted by observing a peculiar property of AG2: certain events do not have any causes. To see why this holds, note that a certain event e is a variable that comes out as true in all complete settings $< M, V >$, and therefore it is impossible to find any V' that is uninformative on e. The same reasoning also shows that certain events c do not have any effects, even if the effect itself is uncertain. Both properties already seem problematic in and of themselves for a definition of causation, but one particularly undesirable consequence is that counterfactual dependence is not sufficient for causation. It can be assumed that AG are unaware of this, for they explicitly subscribe to the sufficiency of counterfactual dependence.[4] Informally, e counterfactually depends on c if both c and e occurred, and if c had not occurred then neither would have e. Formally, this is established by applying our schema (Definition 3.1) with the trivial condition that $M' = M$.

Switch illustrates why counterfactual dependence does not suffice for AG2. Due to the fact that e is certain, it is impossible to find a set V' so that $< M, V' >$ is uninformative on e. Therefore e has no causes according to AG2. Yet e does counterfactually depend on d: setting d to 0 results in a model that satisfies $\neg e$.

To see why, we first consider AG's usual V-representation $< \{e = d \lor b, d = c, b = \neg c\}, \{e, d, \neg b, c\} >$. The intervention that sets d to 0 results in $< \{e = d \lor b, d = 0, b = \neg c\}, \{e, \neg b, c\} >$. Here V does not even satisfy M, because $e \in V$ is inconsistent with the combination of M and $\neg b \in V$, and thus *all formulas are satisfied*, $\neg e$ included. (Recall the definition of satisfaction from Section 3.) We come back to this peculiar behavior in the next section and for now focus on the M-representation of **Switch**, $< \{e = d \lor b, d = c, b = \neg c, c = 1\}, \varnothing >$. Now the result of the intervention is $< \{e = d \lor b, d = 0, b = \neg c, c = 1\}, \varnothing >$, and this model also satisfies $\neg e$.

We can slightly modify AG2 to fix this problem without altering any of AG's desired verdicts for the examples they discuss. The idea is to split up being uninformative on c and e into two steps: first, we simply *ensure* that the resulting model is uninformative on c by removing its equation from M (if it has one) and by removing it from V (if it was in there), and only then do we apply condition C2. (Note that if one uses V-representations, as AG do, this change only affects candidate causes that are not root variables. Presumably the fact that AG exclusively consider root variables as causes in all of their papers explains why this problem went by unnoticed.)

With this modification, counterfactual dependence *is* sufficient for AG2. Say $< M^*, V^* >$ is the result of the first step. Then choosing $V' = V^*$ does the job. If e counterfactually depends on c, then $< M, V >$ satisfies c and e and $< M_{\neg c}, V_{\neg c} >$ satisfies $\neg e$. From the former, it follows that $< M^*_c, V^*_c >$ satisfies e. From the latter, since $M^*_{\neg c} = M_{\neg c}$ and $V^* = V_{\neg c}$, it follows that $< M^*_{\neg c}, V^*_{\neg c} >$ satisfies $\neg e$. This implies that $< M^*, V^* >$ is uninformative on e (in addition to being uninformative on c). Combining all these claims, the result follows.

4.2 Unacceptable Counterexamples

Here is one example of a causal statement which does not leave room for disagreement: for x to cause y, x has to be an ancestor of y in the causal graph that represents the causal dependencies of a causal model. (Causal graphs are standardly used to represent the causal dependencies of variables. In the case of neuron diagrams, the causal graph is obtained by viewing all edges as regular directed edges, independent of whether it contains an arrowhead or a circlehead.) Now consider again the last paragraph preceding Section 4, but instead of evaluating whether or not c causes e in **Early Preemption**, we want to evaluate whether or not c causes a. The same reasoning applies as with e, and we get the unacceptable verdict that c causes a...

To get a sense of how absurd this result is let us apply it to a famous example that AG discuss (Andreas and Günther, 2021a, p. 14):[5]

> Billy and Suzy are throwing rocks at a bottle. ... Suzy throws her rock (c) an instant earlier than Billy does (a). Suzy's rock hits the bottle (d), and so the bottle shatters (e). The shattering of the bottle prevents Billy's rock from hitting the bottle ($\neg b$). The occurrence of the effect e cuts off the backup process started by a.

AG represent this example very similarly to **Early Preemption** (as does everyone else in the literature), and as a result the above causal verdict applies here as well, which means that according to AG2 *Suzy's throw caused Billy's throw*, even though their throws occured entirely independently of each other.[6]

What went wrong here? The problem arises due to the fact that in AG's causal models interventions allow for **backtracking**. Here is an illustration. Say the equation for b is $b = a$, and assume $b \in V$ whereas $a \notin V$. By reading $b = a$ from left to right, we can infer that a holds. This inference is backtracking because it reasons from an effect to its causes. If we now set b to 0 by an intervention, we replace $b = a$ with $b = 0$, and in so doing we block this backtracking inference. More generally, backtracking occurs whenever a variable x is used to infer the values of a variable that is not a descendant of x. This is what happened in **Early Preemption**: before the intervention, the value of a was still open, after the intervention that set $c = 0$ we inferred that $a = 0$, and yet a is not a descendant of c. In standard causal models all interventions block such backtracking inferences, and thus this problem cannot arise.[7]

There are two strategies for AG that suggest themselves to avoid these kinds of unacceptable counterexamples. The first strategy is to redefine an intervention so that it is always non-backtracking, bringing their definition of causal models closer to the standard definition. The second strategy is to leave interventions as they are and enforce a more stringent uninformativeness condition in C2: if we can ensure that the set V' we used above no longer meets C2 when applied to c and a, we obtain that c no longer causes a. The remainder of this paper is dedicated to showing that both strategies lead to dead ends, because they each give rise to novel unacceptable problems.

5 Two Strategies

5.1 Reformulating Condition C2

The second strategy takes inspiration from another one of AG's definitions of causation, AG3 (Andreas and Günther, 2021b). AG3 is formulated using standard causal models that are extended with epistemic concepts such as *being agnostic*, which serves the same purpose as uninformativeness. Interestingly, AG3 contains a different uninformativeness condition than AG2, namely that the model is uninformative on the formula $c \vee e$ (instead of on c and e separately, see Definition 3.2). (Other than that the definitions are quite dissimilar.)

Even though using this condition would be to no avail for our example, as $< M, V' >$ is also uninformative on $c \vee a$ (since this formula is true in V_1 and V_2 but false in V_3), it does open the door to finding a different uninformativeness condition. In particular, we need a condition such that it is satisfied for c and e in **Early Preemption** (so that we keep c causing e), and is not satisfied for c and a (so that we lose c causing a). Looking at V_1, V_2, and V_3, there is exactly one condition which achieves this: $< M, V' >$ should be uninformative on $c \leftrightarrow e$.

Unfortunately choosing this condition is not an option. Consider $< \{e = c\}, \{e, c\} >$, which is the most trivial example possible of c causing e. Clearly we cannot find any V' here such that the model is uninformative on $c \leftrightarrow e$, and thus we fail to reach the undeniable verdict that c causes e. Therefore the second strategy is off the table, leading us to consider the first strategy.

5.2 Redefining Interventions

Recall that we slightly modified AG's definition of an intervention I to ensure that V is also updated to V_I. This minor modification naturally suggests a more substantial modification that sets us on the right path to following the first strategy of solving the **backtracking** problem. When intervening to set c to 0, we motivated the removal of c from V by the fact that V should be updated to reflect the intended consequences of an intervention on c. But what about the consequences this intervention has for the values of variables other than c? Say we also have the equation $e = c$, and $e \in V$. Then the intervention that sets c to 0 still results in logical explosion, despite our slight modification. On the other hand, if the equation instead were $c = e$, then $e \in V$ does not pose a problem at all, as the intervention on c removes this equation from M.

The difference between both of these scenarios gets at the difference between forwardtracking and backtracking inferences: in the first we use the value of a variable to infer the value of one of its descendants, which is what interventions are supposed to allow us to do, whereas in the second we go from a variable to one of its parents, which is precisely what interventions are supposed to prevent (and do prevent under AG's definition). Therefore a natural proposal to avoid the explosion from the first scenario is to redefine an intervention on c such that we remove all literals containing descendants of c from V, thereby preventing any possible conflict that arises between the pre-intervention literals that were stored in V and the literals that result from forwardtracking inferences after the intervention. Adopting this proposal solves the **backtracking** problem.

Let us illustrate this for the problem case at hand: in **Early Preemption** we inferred $\neg a$ from $\neg c$ by using the fact that $\neg b \in V'$. Under the modified definition that we are considering, the intervention that sets c to 0 forces the removal of $\neg b$ from V', and thus this inference is blocked. The resulting model is uninformative on a, and thus c does not cause a. Unfortunately, in the same stroke we get that neither does c cause e, which is precisely the result that AG wanted to avoid in the first place. Alas, we have again reached a dead end.

6 Conclusion

AG's approach to actual causation using causal models is original in that it explores non-standard definitions of causal models and thereby allows for definitions of causation that are qualitatively distinct from existing definitions in the literature. In the case of AG2, however, this approach fails to deliver an acceptable definition of causation. Although AG have two other proposals to rely on, abandoning AG2 does come at a cost for them. Choosing AG1 means to abandon the counterfactual approach to causation (while still being susceptible to Hitchcock's criticism). Choosing AG3 means to abandon the counterfactual approach *and* to give up on the distinction between **Early Preemption** and

Switch.

It is important to note that the argument in this paper did not proceed as is standard in the literature on causation. Usually one argues against a definition of causation by constructing examples for which the definition offers what are alleged to be counterintuitive verdicts. Such arguments leave room for disagreement, because intuitions vary considerably, and so do views on the importance of intuitions for judging definitions of actual causation (Beckers, 2021a; Glymour et al, 2010). The argument here presented does not leave room for disagreement, for it does not rely on challenging AG's desired causal verdicts on **Early Preemption** and **Switch**, nor does it rely on accepting or challenging intuitions on any other disputable example. Therefore the conclusion is that AG2 is damaged beyond repair, even when judged by AG's own light.

Notes

[1]The example is the result of changing the equations from $d = c$ and $b = \neg c$ to $d = c \wedge f$ and $b = \neg c \wedge g$, where f and g represent some background conditions that are actually fulfilled. This change preserves the symmetry between the upper and the lower structures as well as c's role in acting as a switch between these two structures, and thus one would expect the same causal verdict as for **Switch**.

[2]To be clear, there do exist definitions of causation that are formulated using an *extension* of standard causal models. The most notable example is the use of a normality ranking to deal with the problem of structural isomorphisms (Halpern and Hitchcock, 2010). Beckers and Vennekens (2018) extend standard causal models with temporal information to define causation, although I have

subsequently developed a modification of that definition that no longer does so in Beckers (2021b).

[3]Here we are assuming that the equations are acyclic, which is typical when studying actual causation.

[4]Concretely, they state that "dependence between distinct occurring events is taken to be sufficient for causation" (Andreas and Günther, 2021a, p.1-2). For sake of completeness we add that AG3 also suffers from this problem.

[5]This example is a causation classic in the literature called Late Preemption.

[6]For sake of completeness it should be pointed out that the first causal model AG suggest for modelling Late Preemption is inconsistent. According to their model (which is cyclic), if Suzy had not thrown then there is no solution to the equations. This is clearly unintentional, for it is obvious that if Suzy had not thrown then Billy's rock would have hit the bottle and broken it instead. The representation to which we here refer is a second (and unproblematic) model that AG consider.

[7]An alternative, complementary, description of the problem here is that AG's approach allows for conditioning on a collider: the reason we can infer $a = 0$ is that a and c become dependent once we condition on their common effect b. I thank an anonymous reviewer for this suggestion.

References

Andreas H, Günther M (2020) Causation in terms of production. Philosophical Studies 177(1565-1591)

Andreas H, Günther M (2021a) Difference-making causation. Journal of Philosophy 118:680–701

Andreas H, Günther M (2021b) A ramsey test analysis of causation for causal models. British journal for the philosophy of science 72:587–615

Beckers S (2021a) Causal sufficiency and actual causation. Journal of Philosophical Logic 50:1341–1374

Beckers S (2021b) The counterfactual ness definition of causation. In: Proceedings of The Thirty-Fifth AAAI Conference on Artificial Intelligence, pp 6210–6217

Beckers S, Vennekens J (2017) The transitivity and asymmetry of actual causation. Ergo 4(1):1–27

Beckers S, Vennekens J (2018) A principled approach to defining actual causation. Synthese 195(2):835–862

Glymour C, Danks D, Glymour B, Eberhardt F, Ramsey J, Scheines R, Spirtes P, Teng CM, Zhang J (2010) Actual causation: a stone soup essay. Synthese 2:169–192

Hall N (2007) Structural equations and causation. Philosophical Studies 132(1):109–136

Halpern J, Hitchcock C (2010) Actual causation and the art of modeling. In: Causality, Probability, and Heuristics: A Tribute to Judea Pearl, London: College Publications, pp 383–406

Halpern J, Pearl J (2005) Causes and explanations: A structural-model approach. part I: Causes. The British Journal for the Philosophy of Science 56(4):843–87

Halpern JY (2016) Actual Causality. MIT Press

Hitchcock C (2009) Structural equations and causation: six counterexamples. Philosophical Studies 144:391–401

Lewis D (1973) Causation. Journal of Philosophy 70:113–126

Pearl J (2009) Causality: Models, Reasoning, and Inference; 2nd edition. Cambridge University Press